

APPENDIX D

GLOSSARY

Glossary to the Neural Network Appendix

BackErrorPropagation Method - neural network training method to find the optimum values of the connection strength between hidden nodes and input/output nodes (weights). For each data example there is a forward pass from the input nodes to the output nodes to determine the current neural network's output values, followed by a backward pass from the output nodes to the hidden neurons to determine (based on the difference between the predicted output values and the desired values) how and in which direction the weights need to change. The name *backpropagation* is derived from the process of propagating the error information backward from the output nodes to the hidden nodes (Smith, 1996); also known as the steepest descent method.

Batch Learning - calculation of the residual (i.e., the difference between predicted and measured output value) for each data example within the entire training matrix during feed forward propagation to the neural network, summation of the individual residuals and calculation of an overall average residual which then is fed backward once to adjust the layer weights of the neural network.

Conjugate Gradient Method - neural network training method that performs a search along two directions. This neural network training method generally produces faster convergence than back-error-propagation. It is a second-order weight optimization method that involves calculating an approximation of the second derivative of the error with respect to a weight change and can be computationally expensive.

Coefficient of Determination, R^2 - estimation of the degree of interrelation between two variables in a manner not influenced by measurement units. It is the ratio of the explained variation to the total variation of a population (Spiegel, M.R., 1961).

$$R^2 = \frac{\text{explained variation}}{\text{total variation}} = \frac{\text{explained variation}}{\text{unexplained variation} + \text{explained variation}}$$

$$R^2 = \frac{\sum_j (Y_{est} - \bar{Y})^2}{\sum_j (Y - \bar{Y})^2}$$

In the neural network study, R^2 was used as a linear measure of the correlation between input and output values, reflecting the goodness of the linear regression fit. R^2 values near zero mean almost no linear relation between the variables (pointing to highly non-linear relationship); R^2 values

close to one indicate a highly linear relationship. R^2 is dimensionless and always non-negative.

DAF - (Dilution Attenuation Factor), dimensionless ratio of the leachate concentration to the receptor well concentration. The DAF represents the combined effects of the site characteristics, hydrogeologic settings and chemical-specific parameters on the receptor well concentration.

Data Sample - particular combination of parameters (input/output) with their associated values.

Data Matrix - rectangular array of numbers, or elements, arranged in m rows and n columns, where each row represents one data sample.

EPACMTP - (EPA's Composite Model for Leachate Migration with Transformation Products), probabilistic modeling software used to produce the raw data for the neural network training by simulating the subsurface fate and transport of waste constituents leaching from land disposal units; processes accounted for: advection, hydrodynamic dispersion, linear/non-linear sorption, first-order decay (hydrolysis).

IWEM - (Industrial Waste Management Evaluation Model), windows-based software for determining the groundwater protection afforded by various WMU liner systems. Consists of a two-tiered approach for evaluation of WMU liner systems: Tier I - conservative analysis based on national data; Tier II - analysis with a limited set of site-specific data based on the neural networks described in this document.

Feed Forward Propagation - neural network method that passes values from the input parameter in the input layer to the hidden neuron in the hidden layer and then to the output parameter in the output layer.

Hidden Neuron - a neuron or processing unit within a neural network that connects the known input and output parameters. Hidden neurons contain the sigmoid transfer function needed to calculate the weight adjustment factors; connected by weights with the input and output parameters, also called the hidden node.

Hidden Layer - a layer of neurons or processing units within a neural network that are not directly visible to the user, as the input and output values are. See hidden neuron.

Hlearning Rate - constant coefficient applied during adjustment of the neural network weights between the input and the hidden layer after each batch; also called hidden rate. The Hlearning rate should be twice that of the learning rate (output rate).

Input/Output Value - input value: the independent parameter data values that are used to predict the dependent (output) parameter values; output value: the dependent parameter value predicted by the neural network, based on combinations of input parameter values.

Input Layer - a layer of neurons or nodes that contains all neural network input parameters and their data values. The input layer is linked directly to the hidden layer(s).

Interpolation - estimation of input values x_{ii} between given input values x_i for a known function $f(x_i)$ based on the assumption that in a neighborhood of the x_i -value in question, f can be approximated by a polynomial p , whose value at that x_i is then taken as an approximation of the value of f at that x_i ; types: linear, and quadratic (Kreyszig, 1988).

Interrogation - a test of the predictions of a trained neural network by querying it with desired input values.

Land Application Unit - above ground waste management unit in which a liquid waste or sludge is spread as a thin layer on the ground surface to promote biochemical decay of waste constituents. Assumed maximum lifetime of a land application waste management unit and the resulting leachate pulse duration is 40 years.

Landfill - a permanent waste management unit in which solid waste is deposited generally above the ground surface. The base of the landfill may contain a liner to prevent leakage of waste constituents into the groundwater. The duration of the leaching period for landfills is dependent on the initial amount of contaminant, the infiltration rate, the landfill dimensions, and waste density. For large landfills, the pulse duration can approach steady-state conditions.

LCTV - Leachate Concentration Threshold Value, the maximum allowable leachate concentration of a waste constituent in a waste that is intended to be disposed in an industrial waste management unit.

Learning Rate - rate of neural network learning (ability to predict dependent parameter values based on independent parameter values) measured while adjusting the weights of the connections between the hidden layer and the output layer.

Model Dimension - dimensionality of a neural network model based on one dimension per input/output parameter set. For example, 7 inputs and 1 output results in seven model dimensions.

Neural Network - a predictive model of a real world system in which dependent parameter values

(outputs) are determined based on a set of independent parameter values and the relationships of the connections between these values, which are based on hidden neurons and weights. Weights are adjusted during the development of a neural network to minimize the error between the neural network answer (prediction) and the desired output (Swingler, 1996).

Neural Network Learning/Generalization - learning: the development of a predictive capability of a neural network model based on patterns of data (i.e., input values, layer weights, and hidden neurons); generalization: the ability of a neural network model to predict well over a wide range of input values.

Number of Counts - number of neural network model iterations or adjustments of weights.

Output Layer - layer of neurons or nodes in a neural network that contains all output parameters and their data values.

Overtraining - the quality of a neural network in which it predicts very well for data values on which it was trained, but cannot predict well for input values on which the neural network has not been trained. Indicated when during the neural network training process the error of the validation data set suddenly increases (i.e., R^2 value decreases) when the error is expected to steadily decrease.

Parameter Range - range of input parameter values that encompasses entire set of values which are used in training and validating neural networks; defined by a minimum and maximum value.

Parameter - independent (input) or dependent (output) variable contained in the input or output layer of a neural network.

Random Sample - a data sample created using a random number generator to produce an unbiased data set which can be used to test or train a neural network.

Residual - the difference between an observed (e.g. EPACMTP) output value and the neural network predicted value, also called error.

Response Surface - the output of a model or system as a function of a set of input parameters represented by a line or multi-dimensional surface. For example, a two-dimensional response surface is created by connecting the input data points for each input parameter (1D) and output parameter (1D).

Stratified Sample - data sample produced with pre-determined (non-random) combinations of

input parameter values.

Surface Impoundments - waste management unit in which a liquid waste is held in a pond-like enclosure above or partly integrated in the ground surface. The assumed maximum life of a surface impoundment and its duration of leaching period is 20 years.

Test Set/Master Test Set - an independent, randomly established set of data examples used to test the accuracy of neural network predictions during or after the training process. Master test sets are used to test the predictive capability of a neural network, whereas test sets maybe totally or partially incorporated into a training data set.

Tlearning Rate - initial neural network learning rate (rate at which neural network develops predictive capabilities) for all threshold connections.

Training parameters - settings of the neural network software (NNModel) used to start the neural network training process, including: learning rates, number of hidden neurons, number of training counts, or training method; settings such as number of training counts or number of hidden neurons may be changed during the neural network training process.

Training - process of developing the predictive capability of a neural network consisting of adjusting the weights of a neural network by passing a set of data examples of input-output values through the model and adjusting the weights between the input/output parameters and the hidden neurons to minimize the error between the neural network prediction and the desired answer (Swingler, 1996).

Validation - method used to test the predictive capability of a neural network using independent training data examples and to monitor the progress of the training process; validation data sets can be completely or partially incorporated into a training data matrix to improve the neural network predictive capability.

Variable Space - also called data space; a term used to describe the range of neural network input parameter values and combinations of these values.

Waste Piles - an uncovered temporary waste management unit used to store solid waste above the ground surface, with an assumed maximum lifetime and duration of leaching period of 20 years.

Waste Management Unit - permanent or temporary facility used to store/manage industrial waste. Includes landfills, waste piles, surface impoundments, and land application units.

NOTES